Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

# A multifunnel energy landscape encodes the competing $\alpha$-helix and $\beta$-hairpin conformations for a designed peptide†

Debayan Chakraborty,[a§*] Yassmine Chebaro,[b] and David J. Wales[a*]

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Depending on the amino-acid sequence, as well as the local environment, some peptides have the capability to fold into multiple secondary structures. Conformational switching between such structures is a key element of protein folding and aggregation. Specifically, understanding the molecular mechanism underlying the transition from an $\alpha$-helix to a $\beta$-hairpin is critical because it is thought to be a harbinger of amyloid assembly. In this study, we explore the energy landscape for a 18-residue peptide (DP5), designed by Araki and Tamura to exhibit equal propensities for the $\alpha$-helical and $\beta$-hairpin forms. We find that the degeneracy is encoded in the multifunnel nature of the underlying free energy landscape. In agreement with experiment, we also observe that mutation of tyrosine at position 12 to a serine shifts the equilibrium in favour of the $\alpha$-helix conformation, by altering the landscape topography. The transition from the $\alpha$-helix to the $\beta$-hairpin is a complex step-wise process, and occurs via collapsed coil-like intermediates. Our findings suggest that even a single mutation can tune the emergent features of the landscape, providing an efficient route to protein design. Interestingly, the transition pathways for the conformational switch seem to be minimally perturbed upon mutation, suggesting that there could be universal microscopic features that are conserved among different switch-competent protein sequences.

## Introduction

Anfinsen's thermodynamic hypothesis,[1] propounded more than fifty years back, suggests that the three-dimensional organisation of protein structure is largely dictated by its amino acid sequence. Although this idea remains one of the cornerstones of modern molecular biology, recent discovery of protein 'conformational switches' seems to challenge the notion of a sequence-to-structure paradigm.[2–6] Such structural plasticity usually manifests at the level of secondary structure, with identical or similar sequences adopting distinct folds, depending on the context or environmental factors.[7–10] One of the earliest examples of context-dependent adaptability was provided by Kabsch and Sander through mining a structural database.[11] Subsequently, Minor and Kim showed that an 11-residue fragment within a small protein could adopt either $\alpha$-helix or $\beta$-hairpin conformation depending on its relative position in the primary sequence.[12] In another study, Cordes and coworkers[13] demonstrated that two members of the Cro repressor family, with elevated sequence identity, display a striking fold switch from $\alpha$-helix to $\beta$-sheet in a twenty-five residue segment near their respective C-terminal regions. Aside from naturally occurring sequences, a significant number of peptides and proteins have been designed to meet the Paracelsus Challenge,[14] where the objective is to induce switching behaviour in a globular protein by changing no more than half the sequence. Regan and coworkers[15,16] were the first to demonstrate that a predominantly $\beta$-sheet region within the B1 domain of IgG binding protein G could be transformed to a four-helix bundle conformation via rational design. Subsequently, Bryan, Orban and coworkers successfully designed proteins with 88% sequence identity, but different monomeric folds.[17,18] These experimental successes have been complemented by recent effort towards the *in silico* design of switchable peptides.[19–21]

The transition from an $\alpha$-helix to a $\beta$-sheet conformation is often suggested as a precursor in protein misfolding and aggregation, the underlying cause of many neurological disorders.[22–24] For example, the cellular prion protein primarily consists of helices, whereas the pathological form of the protein responsible for neurodegenerative diseases, such as bovine spongiform encephalopathy and Creutzfeld-Jacod disease in humans, acquires a

[a] *Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW, UK. Email: dw34@cam.ac.uk*

§ *Present address: Department of Chemistry, The University of Texas at Austin, 24th Street Stop A5300, Austin TX 78712, USA. Email: debayan.chakraborty@utexas.edu*

[b] *Department of Integrative Structural Biology, Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC), CNRS UMR 7104, INSERM U964, Université de Strasbourg, 67404 Illkirch, France.*

‡ These authors contributed equally.

† Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/b000000x/

high $\beta$-sheet content.[25] The $\alpha \rightarrow \beta$ conformational switch could also lead to functional structures, and is a crucial step in the folding of certain proteins, such as src SH3[26] and $\beta$-lactoglobulin.[27] Due to its key role in protein folding and aggregation, there has been a significant interest in decoding the key aspects of the $\alpha \rightarrow \beta$ transition, from experiments[28,29] as well as computer simulations.[30–39] Nonetheless, lack of sufficient microscopic insight, in terms of both transition pathways and rates, has precluded a complete understanding of this important conformational switch at the molecular level.

In the present work, we investigate the $\alpha \rightarrow \beta$ conformational switch in the context of an 18-residue peptide designed by Araki and Tamura.[40] Starting with an 11-residue segment from human $\alpha$-lactalbumin, which exists as an $\alpha$-helix in acidic conditions,[41] the authors attempted to induce switching behaviour by adding extra residues to the C-terminal. Based on NOESY spectra, they concluded that one of the designed peptides (DP5), with the sequence INYWLAHAKAGYIVHWTA, exhibits nearly equal propensity to form $\alpha$-helix and $\beta$-hairpin structures. Furthermore, the mutation of the tyrosine at position 12 in DP5 to a serine (denoted as DP3) shifts the equilibrium completely in favour of the $\alpha$-helix.

Several computational studies have focused on understanding the transformation between the two prominent conformational states of DP5 using different enhanced sampling techniques. Okamoto and coworkers employed generalised ensemble molecular dynamics, exploiting the multicanonical-multioverlap algorithm,[42] to characterise the transitions between the $\alpha$-helix and $\beta$-hairpin conformations.[43] Based on a two-dimensional free energy landscape, multiple local minima and putative transition states were characterised, having either partial helix or hairpin structures. It was suggested that the conformational switch to the $\beta$-hairpin conformation is triggered by an initial unwinding of the $\alpha$-helix near the N-terminus. In subsequent work,[44] Okumura and Itoh investigated the transformation pathways of DP5 using the 'helix strand replica exchange method', which shares the same formalism as Hamiltonian replica exchange. To facilitate the conformational switch, umbrella potentials were applied in the dihedral angle space. The authors demonstrated that this approach is more efficient at exploring the conformational space of DP5 than temperature-based replica exchange. Furthermore, the free energy difference between $\alpha$-helix and $\beta$-hairpin ensembles was shown to be approximately zero, in agreement with the experimental results.[40] In another study, Mou and coworkers[45] used the Wang-Landau algorithm in conjunction with a coarse-grained peptide model to map out the free energy landscapes for the DP5 and DP3 sequences. Their work suggests that the interplay of dipole-dipole and hydrogen-bonding interactions plays a key role in regulating the conformational switch, and the degeneracy of the native state. Despite the simplicity of the peptide model employed, the study of Mou and coworkers corroborated a key experimental finding: the mutation of tyrosine at position 12 (DP5) to serine (DP3) lifts the degeneracy, and shifts the balance in favour of the $\alpha$-helix. Interestingly, the $\beta$-hairpin conformation for DP3, though not detected experimentally, appeared as a high-lying minimum on the free energy landscape.

In the present contribution, we use the discrete path sampling (DPS) technique[46,47] to characterise the underlying energy landscapes for the DP5 and DP3 peptides. Within the DPS framework, transition pathways are described geometrically in terms of interconnected minimum-transition state-minimum triples on the underlying potential energy landscape, and an *a priori* choice of reaction coordinates is unnecessary. Our study complements previous work based on the finite temperature string,[36,48] integrated and tempering[49] techniques that also provide a reaction coordinate free approach to investigate the $\alpha \rightarrow \beta$ transition.

We find that the quasi-degeneracy of the native state for the DP5 sequence is encoded by the multifunnel nature of the corresponding free energy landscape, and the coexistence of the $\alpha$-helix and $\beta$-hairpin conformations leads to a low-temperature peak in the heat capacity profile. The transition from the $\alpha$-helix to the $\beta$-hairpin occurs via collapsed coil-like intermediates. Mutation of tyrosine at position 12 to serine (DP3 sequence) reshapes the energy landscape, and lifts the degeneracy, in agreement with experiment. Nonetheless, the molecular mechanism underlying the $\alpha$-helix to $\beta$-hairpin transition remains largely unaltered, and bears close resemblance to those of other switchable proteins,[31,33,50] suggesting that there could be universal features that are conserved across different sequences.

## Methodology

The initial coordinates for the $\alpha$-helix and $\beta$-hairpin conformations of DP5 were taken from previously reported NMR structures[40] (PDB IDs: 2DX3 and 2DX4 for the helix and hairpin respectively). In experiments,[40] the helix and hairpin states were found to coexist at a pH of 4.5. Following previous work[43,44] the peptide terminals were left uncapped, and the histidines were protonated to simulate the acidic conditions employed in the experiments. The initial structure for the DP3 sequence was prepared by mutating the Tyrosine-12 residue in the DP5 structure to a Serine using the mutagenesis plugin available in PyMol.[51]

### Force Field and Solvent Model

The peptides were modelled using a properly symmetrised version of the AMBER99SB force-field.[52] In its original form the AMBER force-field exhibits broken symmetry, caused by some of the improper torsion angles. To restore the symmetry of the potential energy function we adopted the strategy described in previous work by Malolepsza *et al.*[53] To make the landscape exploration efficient, and avoid potentially unimportant configurations resulting from minor rearrangements of the water structure, the solvent effects were treated implicitly using a generalised Born solvent model.[54,55] An effective salt concentration of 100 mM was maintained using the Debye-Hückel approximation.

### Molecular Dynamics Simulations in Implicit Solvent

The molecular dynamics simulations were carried out using the GPU enabled version of the AMBER12 package.[56] No cutoff was employed for the non-bonded interactions, and the simulations were carried out without using periodic boundary conditions. The temperature was maintained at 300 K by coupling the system to a Langevin thermostat, using a collision frequency of 1 $ps^{-1}$.

All simulations were 100 ns in length. Snapshots from the MD trajectories were saved every 10 ps and were locally minimised using the GMIN program.[57]

The DSSP algorithm, available within the *ptraj* module of AmberTools, was used to classify the secondary structure corresponding to each residue of the DP3 and DP5 peptide sequences.

### Landscape Exploration using Discrete Path Sampling

The energy landscapes of the DP3 and DP5 peptides were mapped out using the Discrete Path Sampling (DPS) technique[46,47]. DPS is complementary to techniques based on explicit dynamics, and provides a framework to describe the underlying landscape in terms of databases of stationary points (minima and the transition states that connect them). As the stationary points are located using geometry optimisation in a time-independent fashion, DPS is particularly efficient in probing conformational transitions that occur over a wide array of time scales. It has been used to study 'rare event' dynamics in a diverse range of contexts, from protein and RNA folding[58–60] to cluster rearrangements and polymorphism,[61,62] as well as self-assembly.[63] In this section we briefly discuss the key steps of DPS, and refer readers to earlier work,[46,47] which provides the detailed formalism.

Within the DPS framework, the connectivity between different endpoints (termed as reactant and product states) on the potential energy landscape is described in terms of discrete paths, which consist of a sequence of minima linked by intervening transition states. A stationary point having a single imaginary frequency is identified as a transition state, from the Murrell-Laidler definition.[64] Approximate steepest-descent paths directed parallel and anti-parallel to the eigenvector corresponding to the imaginary frequency terminate at the adjoining minima.[65]

As described in the previous section, an initial sample of minima was obtained by the quenching of snapshots from the molecular dynamics trajectories. DPS runs were carried out to connect different local minima corresponding to the $\alpha$-helix and $\beta$-hairpin conformations, respectively. The doubly-nudged[66] elastic band[67,68] method was used to find initial guesses for transition states between pairs of local minima, starting from an image distribution obtained via the quasi-continuous interpolation scheme.[69] This method exploits the connectivity of the covalently bonded network and prevents unphysical chain crossings in the pathway images. The transition state candidates obtained from DNEB were accurately refined using the hybrid eigenvector-following scheme,[70] until the root-mean-square (RMS) gradient fellow below $10^{-6}$ kcal mol$^{-1}$ Å$^{-1}$. The OPTIM code[71] interfaced with the AMBER9 package[72] was used for all the local minimisations and transition state searches. The geometry optimisations were carried out using a modified version of the L-BFGS algorithm.[73]

The initial discrete path obtained between endpoints of interest is usually kinetically unimportant as it tends to be long, and may have high intervening barriers. To locate more relevant pathways, the stationary point databases were further expanded using various refinement schemes. In particular, the SHORTCUT BARRIER and SHORTCUT schemes, described in previous work[74,75] are quite efficient at identifying pathways characterised by lower energy barriers, and shorter path lengths, respectively. However, spurious frustration may be introduced into the stationary point databases due to undersampling of certain regions of the landscape, and manifests in the form of low-lying minima separated from the product region by high energy barriers. The UNTRAP scheme,[74] which selects minima for reconnection attempts based on the ratio of the potential energy barrier to the potential energy difference to the product region, is used to remove spurious frustration. The databases were systematically expanded by sequential applications of these three schemes until no further changes were observed in terms of path lengths and barrier heights for the $\alpha$-helix to $\beta$-hairpin transition pathways.

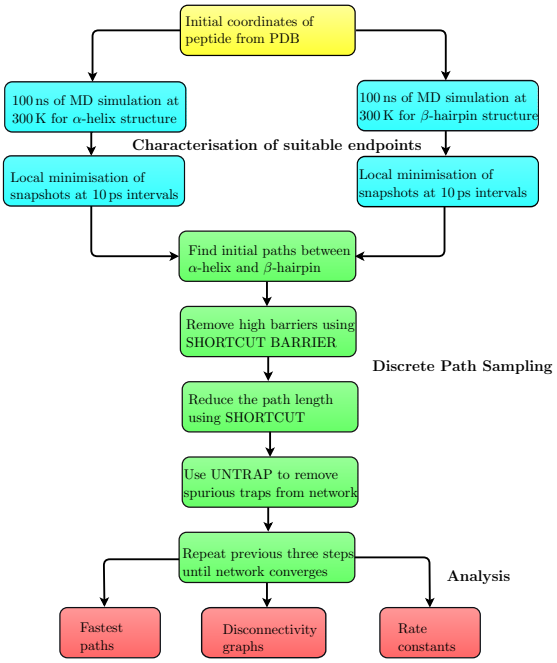


**Fig. 1** A flowchart summarising the key details of the computational methodology.

The rate constant $k^{SS}_{\beta\alpha}$ for the $\alpha$-helix to $\beta$-hairpin transition can be expressed as an infinite sum over discrete paths when the intervening minima are treated within the steady-state approximation, and the dynamics between adjoining minima are assumed to be Markovian.[46,47] The infinite sum is weighted by the occupation probability of the reactant minimum as well as the relevant branching probabilities. The product of the branching probabilities defines the statistical weight of each discrete path.[46,47] We used Dijkstra's shortest path algorithm,[76] with edge-weights corresponding to the product of the branching probabilities[58] to extract the $\alpha$-helix to $\beta$-hairpin transition path that contributes most to the overall rate constant. The flowchart in Figure 1 summarises the key details of the computational methodology.

### Analysis of Free Energies and Global Kinetics

A harmonic approximation[77–79] was employed to estimate the vibrational partition functions associated with the minima and transition states in the stationary point databases. The canonical partition function for local minimum $i$ is expressed as:

$$Z_i(T) = \frac{n_i e^{-U_i/k_B T}}{(h\bar{\nu}_i/k_B T)^{\kappa}} \quad (1)$$

In equation (1), $U_i$ denotes the potential energy of minimum $i$, $n_i$ is the number of distinct permutation-inversion isomers of $i$, $\bar{\nu}_i$ denotes the geometric mean of the normal mode frequencies associated with minimum $i$, and $\kappa$ represents the number of vibrational degrees of freedom. The overall canonical partition function is written as a sum of contributions from the catchment basin of each local minimum.

$$Z(T) = \sum_i^M Z_i(T) \quad (2)$$

Here, $M$ is the number of minima present in the stationary point database. The partition functions for the transition states are defined in the same way as in equations (1) and (2), but the normal mode corresponding to the imaginary frequency is omitted from those expressions. The occupation probabilities and free energies were calculated from the canonical partition function using equilibrium statistical mechanics:

$$F_i(T) = -k_B T \ln Z_i(T), \quad (3)$$

and

$$p_i^{eq}(T) = \frac{Z_i(T)}{Z(T)}. \quad (4)$$

The heat capacity, $C_v$ can be expressed in terms of the partition function, $Z(T)$ using standard thermodynamic relations:

$$C_v = \left(\frac{\partial U(T)}{\partial T}\right)_{N,V}, \quad (5)$$

where $U(T)$ is the internal energy. Using Eq. (1), $C_v$ corresponding to the superposition partition function can be written as:

$$C_v = \kappa k_B - \frac{z_1(T)^2}{k_B T^2 z_0(T)^2} + \frac{z_2(T)}{k_B T^2 z_0(T)} \quad (6)$$

where

$$z_r(T) = \sum_i n_i (V_i)^r \left(\frac{k_B T}{h\bar{\nu}_i}\right)^{\kappa} e^{-U_i/k_B T} \quad (7)$$

The unimolecular rate constant $k_i^{\dagger}(T)$ for minimum $i$ crossing the transition state † at a temperature $T$ is estimated using harmonic transition state theory (TST), as:

$$k_i^{\dagger}(T) = \frac{k_B T}{h} \frac{Z^{\dagger}(T)}{Z_i(T)} \quad (8)$$

In Eq. (8), $Z^{\dagger}(T)$ and $Z_i(T)$ denote the canonical partition functions corresponding to minimum $i$ and transition state †, respectively. Summation of all the $k_i^{\dagger}(T)$ values for all transition states that connect minima $i$ and $j$ gives the total TST rate constant for the transition. The equilibrium occupation probabilities and the TST rate constants obtained using Eq. (4) and Eq. (8), corresponding to all the elementary transitions along the discrete paths, were subsequently used to extract the overall rate constant for the $\alpha$-helix to $\beta$-hairpin transition from the corresponding kinetic transition network using the new graph transformation (NGT) method.[80] To alleviate any bias in the estimate of rate constants that may arise due the original choice of endpoints, a self-consistent lumping scheme[58] was employed. This approach exploits the separation of time scales between the overall conformational transition, and local equilibration within the product and reactant regions. In the regrouping scheme, structures that are separated by free energy barriers below a certain threshold are grouped into one macrostate. The corresponding rate constants are then described in terms of transitions between ensembles, rather than individual minima, thereby making direct comparisons with experiments feasible.[58,81]

To quantify the complexity of the landscape, we estimate the frustration index $\widetilde{f}(T)$ as a function of temperature.[82]. This metric is a quantitative measure of how efficiently a system relaxes to its lowest-energy structure. Self-organising systems are characterised by a single-funnelled landscape with low barriers, and exhibit low levels of frustration, $\widetilde{f}(T)$. In contrast, glassy systems, which are often synonymous with dynamical arrest, have a highly frustrated energy landscape, and may never relax to their lowest energy structure on the observation time scale. Within the computational energy landscape framework, $\widetilde{f}(T)$ is defined as:

$$\widetilde{f}(T) = \sum_{i \neq \mathrm{GM}} \frac{p_i^{\mathrm{eq}}(T)}{1 - p_{\mathrm{GM}}^{\mathrm{eq}}(T)} \left(\frac{U_i^{\dagger} - U_{\mathrm{GM}}}{U_i - U_{\mathrm{GM}}}\right). \quad (9)$$

In Eq. 9, the summation includes all minima in the database excluding the global minimum, GM, $p_i^{\mathrm{eq}}(T)$ is the temperature dependent equilibrium occupation probability of minimum $i$, $U_i$ is the potential energy of minimum $i$, $U_{GM}$ is the potential energy of the global minimum, and $U_i^{\dagger}$ is the potential energy of the highest transition state on the lowest energy path between minimum $i$ and the global minimum.

### Disconnectivity graphs: visualisation of energy landscapes

The free energy landscapes for the DP3 and DP5 peptide sequences were visualised in terms of disconnectivity graphs.[83–86] This diagrammatic representation of the landscape is powerful yet simple, and represents the barriers between different local minima.[87] A disconnectivity graph segregates the landscape into disjoint sets of local minima known as 'superbasins',[83] at regular intervals of energy. Minima within each superbasin are mutually accessible via transition states that lie below a certain threshold, whereas transitions out of a superbasin must surmount higher energy barriers. Basin analysis is performed at regular energy thresholds to yield the desired resolution. In its original form, the horizontal axis of the disconnectivity graph is arbitrary, and does not correspond to any reaction coordinate.[65]

## Results and Discussion

### The NMR structures exhibit enhanced flexibilities

The helix and hairpin conformations of the DP5 sequence exhibit substantial flexibility on the time scale of the MD simulations, with average backbone RMSD from the corresponding NMR structures being 5.5 Å and 6.9 Å, respectively (supporting information, Figure S1). Snapshots corresponding to the different conformations

identified from the MD simulations are shown in Figure 2.

In the helix conformation exhibiting the lowest RMSD ($\alpha_{\mathrm{RMSD}}$), the secondary structure of residues Y3 to A10 is classified as $\alpha$-helical according to the DSSP criterion. The structure with the lowest RMSD in the hairpin ensemble ($\beta_{\mathrm{RMSD}}$), exhibits turns near residues A6, H7, K9, A10 and G11. Interestingly, none of the residues in $\beta_{\mathrm{RMSD}}$ satisfy the DSSP requirements for a $\beta$-bridge conformation. Although our observation seems counter-intuitive at first glance, it is in line with two previous computational studies[43,49] where it was argued that the NMR description of a $\beta$-hairpin, based on NOE distances, need not be commensurate with the DSSP criterion. As outlined in the methods section, snapshots from the MD trajectories were locally minimised to identify the lowest energy configurations in the helix and hairpin ensembles, respectively. The lowest energy helical conformation ($\alpha_{\mathrm{lowest}}$) does not have a continuous helical region, unlike the $\alpha_{\mathrm{RMSD}}$ structure. Instead, residues Y3 to A6 and I13 to W16 form short helical fragments, which are separated by a turn in the middle of the sequence. The existence of similar low-lying partial helical conformations with a turn region was also predicted by earlier simulations[43,44,49] that employed a different combination of force field and water model. In contrast to the $\beta_{\mathrm{RMSD}}$ conformation, which comprises only turns, the lowest energy hairpin structure ($\beta_{\mathrm{lowest}}$) exhibits an anti-parallel $\beta$-sheet, stabilised by hydrogen-bonding interactions between residues L5–V14, and H7–Y12, respectively. The B factor plots depicted in Figure 3 indicate that different residues contribute to the overall flexibility of the helix and hairpin conformations. The tendency of the long $\alpha$-helix to bend during the course of the MD simulation and form partial helical structures is reflected in the relatively high B factors associated with the residues in the middle of the sequence. In the hairpin conformation, the residues in the middle of the sequence exhibit minimal fluctuations, whereas those closer to the termini appear more mobile.

Although standard molecular dynamics provides insight into conformational dynamics occurring over short time scales, it is prone to kinetic trapping, especially for landscapes featuring broken ergodicity.[65,88] To enhance the sampling for the DP5 sequence, and obtain mechanistic insight into the $\alpha \longleftrightarrow \beta$ conformational switch, DPS simulations were seeded from the structures depicted in Figure 2. After initial paths were characterised between the $\alpha$-helix and $\beta$-hairpin structures, the rest of the local minima constituting the helix and hairpin ensembles identified from the MD simulations were systematically added to the stationary point databases (kinetic transition network). This step involved connection-making attempts between the different minima in the helix and hairpin ensembles, and $\alpha_{\mathrm{lowest}}/\beta_{\mathrm{lowest}}$ or $\alpha_{\mathrm{RMSD}}/\beta_{\mathrm{RMSD}}$, whichever is closer to the selected minimum in terms of Euclidean distance. The stationary point databases were further expanded by refining the initial discrete paths using the schemes described in the Methods section, following the procedure outlined in Figure 1. The network is deemed to be converged when the length of the 'fastest path' between the $\alpha$-helix and the $\beta$-hairpin structures, as well as the rate constant for the corresponding transition, remain invariant with respect to the addition of new stationary points. The smallest transition network satisfying this criterion consisted of 50,640 minima and 71,563 transition states.

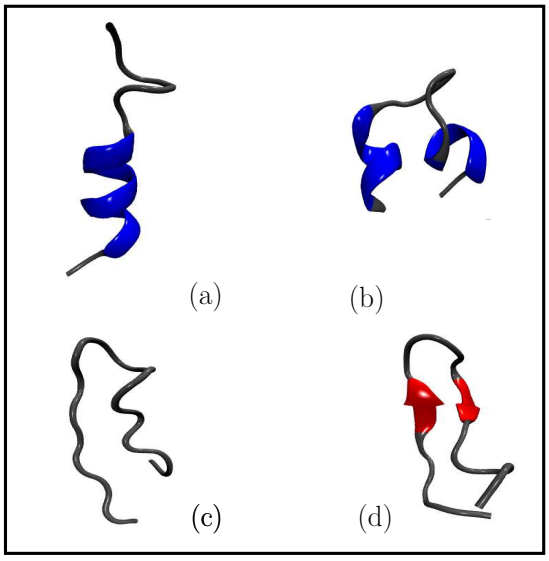


**Fig. 2** Different conformations for the DP5 sequence identified from the MD simulations. (a) The $\alpha_{RMSD}$ conformation, which exhibits a RMSD of 1.3 Å from the NMR structure (2DX3). (b) The lowest energy $\alpha$-helix conformation. (c) The $\beta$-hairpin conformation exhibiting the lowest RMSD from the NMR structure (2DX4). (d) The lowest energy $\beta$-hairpin conformation.

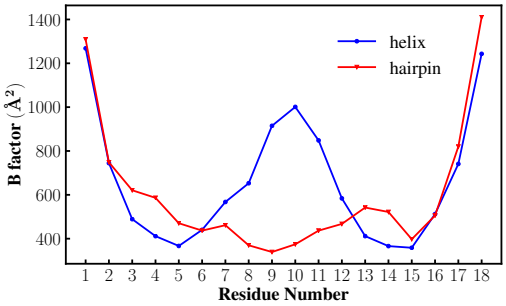


**Fig. 3** The estimated B factors at 300 K for the helix and hairpin conformations.

### A multifunnel energy landscape encodes alternate secondary structures

The free energy landscape computed at 300 K for the DP5 sequence, described in terms of ensemble of stationary points, is depicted in the form of a disconnectivity graph in Figure 3. The branches are coloured according to the type of secondary structure, identified by the DSSP algorithm. Specifically, blue branches lead to minima in which at least six residues are classified as helical; red branches correspond to minima that adopt hairpin-like structures, with at least six residues adopting a $\beta$-bridge orientation; green branches correspond to turn structures that do not have any residue classified as helical or $\beta$-bridge; all other branches are coloured black. Although the segregation based on secondary structure is quite good, it is far from perfect. The intermixing of colours in different regions of the graph indicate that structural

metrics alone are insufficient to faithfully represent the complex features of the landscape.[89] Here, the use of 'secondary structure' is merely to aid visualisation, and we stress that unless a robust kinetic metric, such as the recursive regrouping scheme[58] is used to distinguish different states, a misleading picture of the kinetics may be obtained.[87,90]

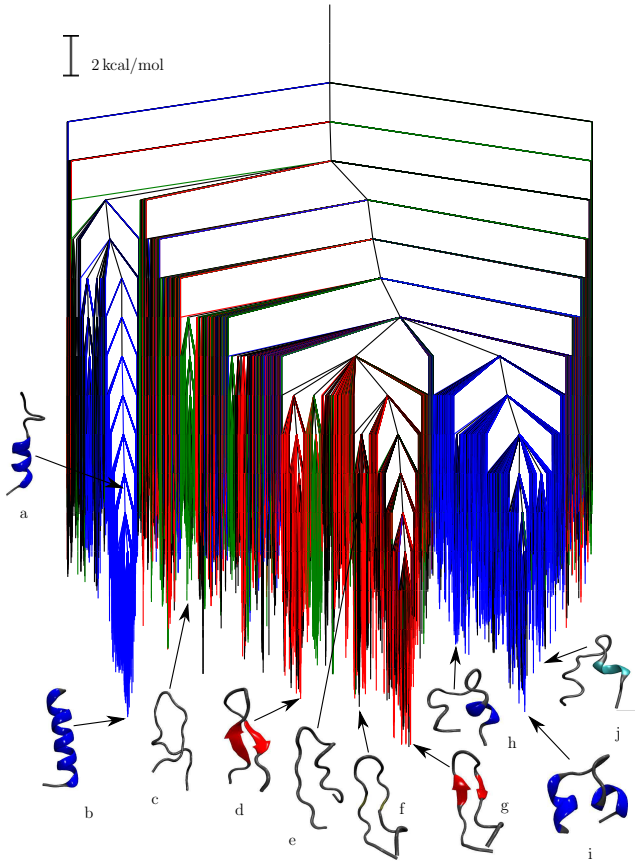


**Fig. 4** The free energy landscape for the DP5 sequence computed at 300 K. Blue branches lead to helical structures, while red branches typically lead to hairpin structures. Some representative snapshots from the different structural ensembles are also shown: $\alpha_{\mathrm{RMSD}}$ (*a*), long $\alpha$-helix (*b*), turn structure (*c*), $\beta$-sheet structure (*d*), $\beta_{\mathrm{RMSD}}$ (*e*), low-energy hairpin without $\beta$-bridge, but consisting of hydrogen bonds (*f*), $\beta_{\mathrm{lowest}}$ (*g*), partial helix with an $\alpha$-helix at the N-terminus (*h*), partial helix with $\alpha$-helices at both termini (i), partial helix with a $3_{10}$-helix at the N-terminus (j). The different secondary structures are coloured according to the scheme described in Figure 2.

The landscape exhibits multifunnel character, with the low-lying region dominated by ensembles corresponding to partial helices, long $\alpha$-helix, hairpins with $\beta$-bridges, and structures exhibiting turns. Snapshots of peptide conformations constituting the different ensembles are shown superimposed on the disconnectivity graph (Figure 4). The lowest energy $\alpha$-helix (snapshot *i*) and $\beta$-hairpin (snapshot *g*) structures identified from MD simulations lie at the bottom of the major blue and red funnels, respectively. The ensembles populating the two funnels are structurally heterogeneous. In addition to conformations similar to $\alpha_{\mathrm{lowest}}$, the partial helix funnel also contains structures exhibiting either an $\alpha$-helix or a $3_{10}$-helix exclusively at the N-terminus, $\alpha$-helix or $3_{10}$-helix exclusively at the C-terminus, $3_{10}$ helices at both termini, or an $\alpha$-helix at one terminus and a $3_{10}$-helix at the other. In Figure 4, we show two of the variants (snapshots *h* and *j*), and the rest are included in the supporting information (Figure S2). Within the red funnel comprising the $\beta_{\mathrm{lowest}}$ structure (snapshot *g*), we identified several hairpins that do not have any $\beta$-bridges, but exhibit turn regions (snapshot *f*). Unlike the turn structures denoted by the green branches (snapshot *c*), these hairpins retain nearly all the canonical hydrogen-bonds that are present in $\beta_{\mathrm{lowest}}$ (supporting information, Figure S3). Conformations similar to $\beta_{\mathrm{RMSD}}$ (snapshot *e*) populate the top of the hairpin funnel. The landscape also features a prominent subfunnel, populated by hairpins exhibiting an alternative $\beta$-sheet structure (snapshot *d*). Here, the $\beta$-sheet is downshifted by two residues and twisted out-of-plane. Compared to $\beta_{\mathrm{lowest}}$, the $\beta$-sheet is slightly longer, extending across three residues on either strand. Long $\alpha$-helix structures constitute the bottom of the narrow blue funnel. Unlike the major funnels, the structural variation near the funnel bottom is minimal, with nearly all helices exhibiting only minor differences in the internal degrees of freedom. The $\alpha_{\mathrm{RMSD}}$ structure (snapshot *a*) is destabilised with respect to the long helix, and lies at the top of the funnel.

The organisation of the landscape is largely consistent with the free energy surfaces obtained from thermodynamic sampling in previous work.[43–45,49] Based on the relative equilibrium populations, these studies concluded that the long $\alpha$-helix structure is destabilised with respect to the hairpin and the partial helical structures. In contrast, we predict the long $\alpha$-helix to be a competing structure on the landscape, separated from the hairpin and partial helices by large barriers. The shape of the landscape in the vicinity of the long helix compared to the organisation near the partial helix and hairpin conformations hints at a possible cause for this discrepancy. The narrowness of the funnel leading to the long helix may make it kinetically inaccessible from the denatured state, which is substantially populated in previous simulations due to the relatively high temperatures employed in thermodynamic sampling. On the other hand, relaxation to the partial helical and hairpin states is likely to be more favourable upon cooling due to the larger basins of attraction associated with the corresponding minima. A similar situation is often encountered in simulations of atomic clusters exhibiting competing morphologies, and has been extensively studied using the landscape framework.[85,91] The topography of the landscape further indicates that during dynamical simulations initiated from the NMR-like helix conformation, the system has a high probability of escaping the basin of attraction of the long helix, and would then evolve towards metastable states that resemble the partial helices in the major blue funnel. This dynamical feature is therefore an emergent property of the landscape, and is accurately captured by our initial MD simulations.

The free energy difference between the long $\alpha$-helix and the the lowest energy $\beta$-hairpin conformation is approximately 0.4 kcal/mol ($\approx 0.6\,k_BT$), consistent with the coexistence of these two folds observed experimentally.[40] Previous simulations also reported a negligible difference in free energy between the $\alpha$-helix and $\beta$-hairpin conformations.[43,45] Local equilibration within the helix and hairpins is fast compared to the $\alpha \rightarrow \beta$ transition, which is associated with a rate constant of $2.7 \times 10^{-9}\,\mathrm{s}^{-1}$, obtained with a regrouping threshold[58] of 3.0 kcal/mol at 300 K. Thus, the two

quasi-degenerate conformations of DP5 are stable on experimental observation time scales. As discussed above, the partial helical basin consists of structures that were previously found using thermodynamic sampling,[43,44] and were also visited on the time scale of our MD simulations. If the partial helical state is selected as representative of the helix ensemble, then the free energy difference between the helix and hairpin conformations is estimated to be 0.9 kcal/mol ($\approx$1.5 $k_BT$). However, the transition between the partial helix and the $\beta$-hairpin is predicted to be faster, with an associated rate constant of $0.45 \times 10^{-1}$ s$^{-1}$ at 300 K.

The transformation from the long $\alpha$-helix to the $\beta$ hairpin requires rather complex structural reorganisation, and is described here in terms of the 'fastest path' in the transition network (Figure 5). The unfolding of the helix starts from the C-terminal region, with the concomitant loss of hydrogen-bonding interactions between residues that are separated by four positions along the chain, namely A8-Y12, K9-I13, A10-V14 and I13-T17. Next, the N-terminal region unfolds, with disruption of hydrogen-bonding interactions between Y3-H7, and L5-K9. The loss of key hydrogen-bonding interactions causes the helix to bend substantially. At this stage, the middle of the sequence still retains some $\alpha$-helical character, and a $3_{10}$ helix is formed near the C-terminal region, between residues Y12 and V14. Subsequently, the $\alpha$-helix in the middle unfolds to form a turn region, and a new hydrogen-bonding interaction results between the distant residues Y3 and H15. This helical structure, with a turn region in the middle, is similar to the conformations that constitute the partial helical basin (Figure 4). Unwinding of the residual helical fragments leads to collapsed coil intermediates exhibiting mostly turn regions. The next stage of the transformation is characterised by the formation of the hydrophobic core of the hairpin, and a subsequent conformational search for the native-like contacts. The formation of the first native hydrogen-bond between L5 and V14 drives the zippering of the rest of the hairpin stem. The final phase is primarily dominated by internal reorganisation of the $\beta$-bridges to native-like conformations. Overall, the pathway for the $\alpha \rightarrow \beta$ transition is in accord with previous findings based on complementary simulation techniques.[31,33,36,38,44]

The conformational transition from the partial $\alpha$-helix structure to the $\beta$ hairpin involves shorter paths, characterised by lower potential energy barriers. The $\alpha$-helical fragment at the C-terminus unwinds first, to form a $3_{10}$ helix. Subsequently, the $\alpha$-helix near the N-terminus unfolds completely. The random coil intermediate formed en route to the $\beta$-hairpin exhibits near perfect alignment of the opposing strands, and is predisposed to form the native hairpin contacts in a rapid downhill fashion. This rapid rearrangement to the hairpin structure is in stark contrast to the multiple collapse and expansion stages of random coil intermediates in pathways originating at the full $\alpha$-helix.

### Reshaping the landscape via mutation

Mutating the tyrosine at position 12 to serine (DP3 sequence) has a pronounced effect on the organisation of the free energy landscape (Figure 6). The corresponding kinetic transition network consists of 48,383 minima and 68,038 transition states. The degeneracy between the $\alpha$-helix and $\beta$ structures in DP5 is absent. The free energy global minimum at 300 K is a long $\alpha$-helix structure (snapshot *h*), which is stabilised with respect to the lowest energy $\beta$-hairpin structure (snapshot *d*) by approximately 3.2 kcal/mol ($\sim$ 5.3 $k_BT$). An array of hydrogen-bonds between residues separated by four positions along the sequence stabilises the long $\alpha$-helix structure. The mutation, in fact, seems to perturb the hydrogen-bonding pattern minimally, as the helix retains most of the interactions found in the DP5 sequence (Figure S4). In contrast, the hydrogen-bonds stabilising the hairpin conformation for DP5 are lost upon mutation (Figure S5), resulting in a deformed $\beta$-sheet structure (snapshot *d*) for DP3. As is evident from Figure 6, the $\beta$-hairpin conformations are not only thermodynamically unfavourable relative to the full helix structure, but are also kinetically inaccessible from the full helix due to the large intervening free energy barriers separating the two funnels. Using a regrouping threshold[58] of 3.0 kcal/mol, the rate constant for the transition from the full $\alpha$-helix to the $\beta$-hairpin is estimated to be $7.8 \times 10^{-15}$ s$^{-1}$ at 300 K, which is several orders of magnitude lower than that for the DP5 sequence. A combination of thermodynamic and kinetic factors therefore rationalise why the $\beta$-hairpin conformation was not detected, and consequently no conformational switching was observed, on the time scale of the NMR experiment.[40]

The free energy landscape of DP3 also features a diverse range of partial helix structures (snapshots *f*, *i*, and *j*), which exhibit a bend in the middle of the sequence, and either $3_{10}$ or $\alpha$ helices at the termini. However, they are no longer competing structures on the landscape, and exist mostly as high-lying minima at the top of the helix funnel. Interestingly, upon mutation, the overall population of turn structures seems to increase as compared to the DP5 sequence, and a major funnel (denoted by green branches in Figure 6) comprising exclusively turn structures emerges on the landscape. In many of the low-lying turn conformations within the green funnel, the two strands are approximately aligned as for a $\beta$-hairpin structure (snapshot *e*), but the canonical hydrogen-bonding interactions, which provide additional stability, are absent.

The heat capacity profiles, and the frustration index (Figure 7) provide further insight into how mutation reshapes the free energy landscape. For the DP5 sequence, the competition between the $\alpha$-helix and $\beta$-hairpin conformations results in a prominent peak in the low temperature region of the heat capacity curve. Such solid-solid type transitions are reminiscent of cluster isomerizations,[91] in which the interplay between enthalpy and entropy switches the free energy global minimum with temperature. Upon mutation, the low temperature peak disappears, indicating that the mutant sequence, DP3, no longer supports the degeneracy. At high temperatures, unwinding of the $\alpha$-helix near the terminals occurs for both sequences. This transition corresponds to a shoulder in the heat capacity profile for the DP5 sequence. In contrast, for DP3, helix unwinding is associated with a broad peak, indicating that a higher change in internal energy is required for unwinding due to the increased depth of the helix funnel. The $\widetilde{f}(T)$ for the DP5 sequence is at least an order of magnitude higher than the DP3 sequence at lower temperatures, in line with the multifunnel character of the free energy landscape, supporting competing

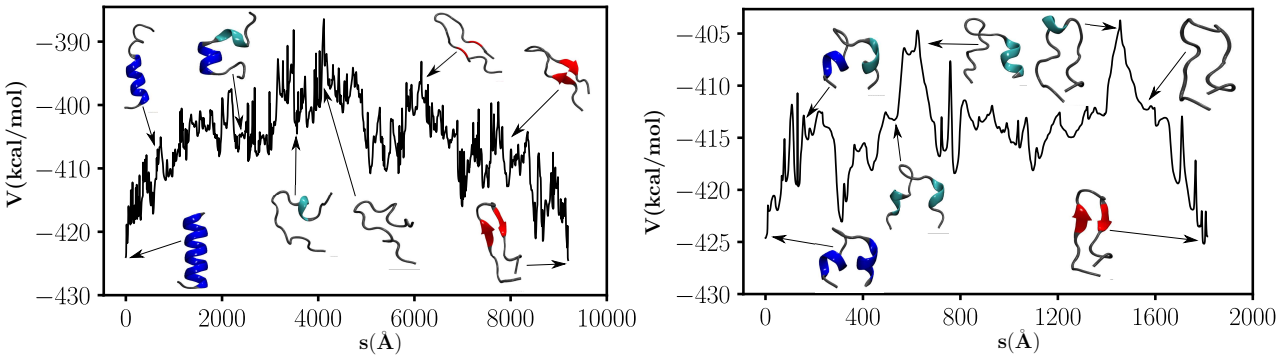


**Fig. 5** Left: The $\alpha$ to $\beta$ transition for the DP5 peptide sequence. The mechanism is described in terms of the path that contributes maximally to the global dynamics. Here, $s$ is the integrated path length. Right:The transition from a partial helix conformation (snapshot $h$ in Figure 4) to the $\beta$ hairpin. Some representative structures encountered along the pathways are shown superimposed.

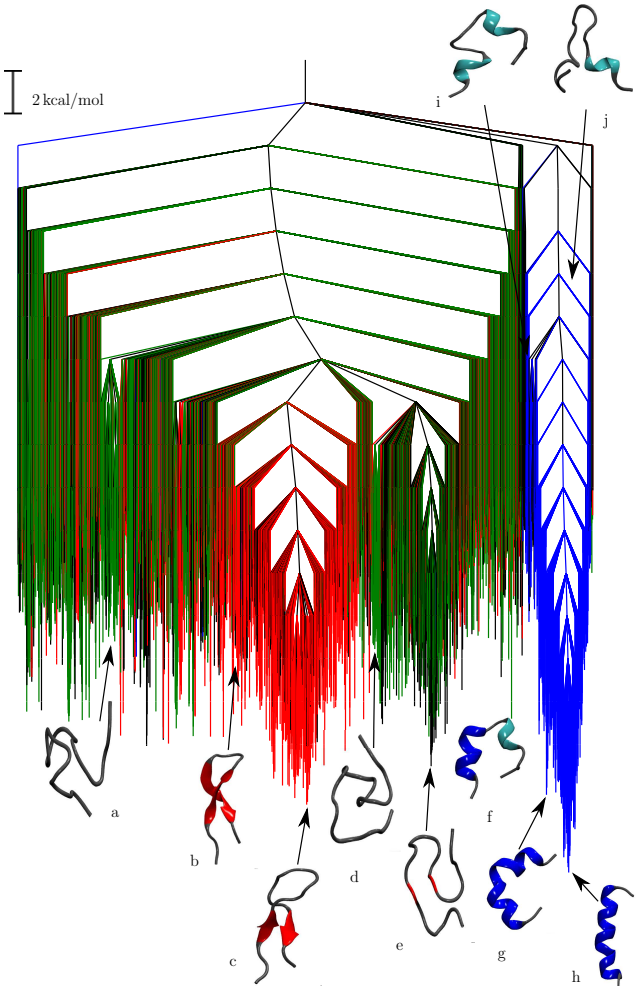


**Fig. 6** The free energy landscape for the DP3 sequence computed at 300 K. The colour coding is same as in Figure 4. Some representative snapshots from the different structural ensembles are also shown: turn structures ($a$, $d$, and $e$), beta hairpin structures ($b$ and $c$), partial helix structures ($f$, $i$ and $j$), a bent helix ($g$), and the full $\alpha$-helix ($h$).

structures. On the other hand, the lower $\widetilde{f}(T)$ for DP3 suggests the multifunnel character is diminished substantially upon mutation, and consequently the quasi-degeneracy between the $\alpha$-helix and $\beta$-hairpin conformations is broken. Our observations recapitulate the findings of previous studies that have attributed multifunctionality and coexistence of different folds to the multifunnel nature of the underlying landscape, and an evolutionary optimisation of frustration.[50,92–94]

Despite the loss of degeneracy upon mutation, it is still instructive to assess the microscopic mechanism underlying the transformation between the full $\alpha$-helix and the $\beta$ conformation (Figure 8). As expected, the intervening potential energy barriers are somewhat higher compared to the DP5 sequence (Figure S6). In the early phase of the transition pathway, the helix bends in the middle to form structures similar to snapshot $g$, which lies near the bottom of the helix funnel (Figure 6). This step is followed by the unwinding of the $\alpha$-helix to form $3_{10}$ helices first near the N-terminus, and then near the C-terminus, resulting in conformations similar to snapshots $f$ and $i$ shown in Figure 6. Subsequently, all helical fragments are broken to form collapsed coil-like structures, exhibiting mostly turn regions. The middle of the transition pathway is dominated mostly by conformational fluctuations of these structures and the search for $\beta$-hairpin type contacts. During the last stages of the pathway, the strands approximately align, and contacts are established between residues H7 and S12, and L5 and H15 to form a high energy $\beta$ hairpin structure. Finally, the L5-H15 contact is broken, and the $\beta$-sheet downshifts to form the low-energy $\beta$-hairpin (snapshot $c$ in Figure 6). Interestingly, the alteration of the landscape topography upon mutation does not perturb the transition mechanism appreciably.

In light of the present study, and several others on proteins of varying complexity,[33,50,95,96] it seems that despite sequence-specific effects, there are common themes underlying the $\alpha$ to $\beta$ transition mechanism, where insight at the molecular level may be crucial for decoding the key aspects of protein folding and aggregation.

## Conclusion

In this work, we have shown that a multifunnelled free energy landscape encodes quasi-degeneracy between the $\alpha$-helix and

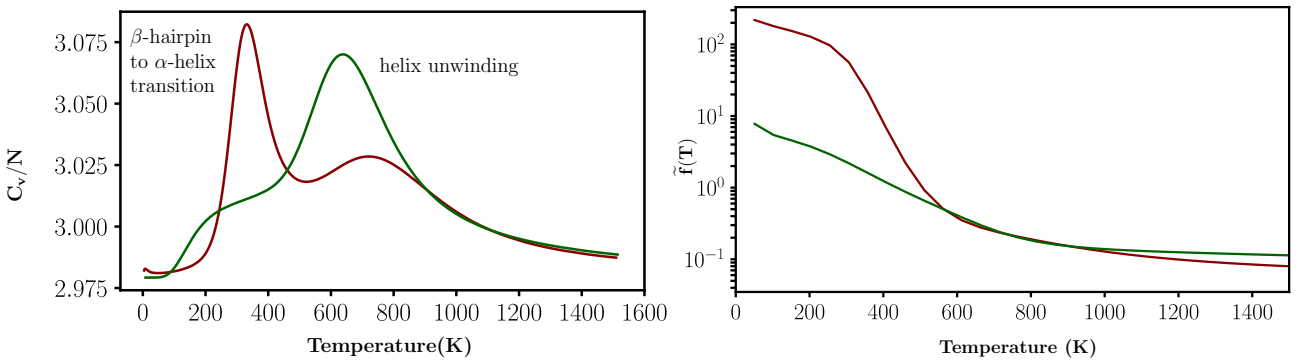


**Fig. 7** Left: The normalised heat capacity profile, $C_v/N$ computed from the database of minima for the DP5 sequence (red), and DP3 sequence (green). Right: Frustration index, $\widetilde{f}(T)$, as a function of temperature for the DP5 sequence (red) and the DP3 sequence (green).

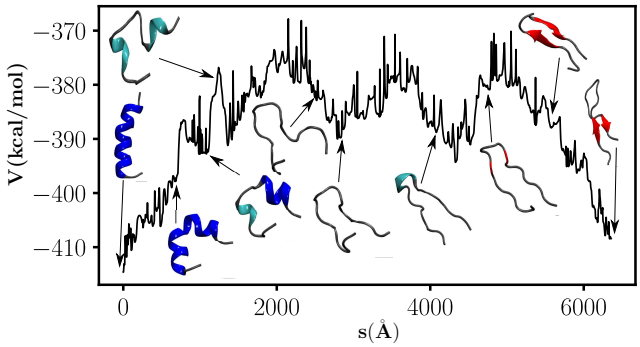


**Fig. 8** The $\alpha$ to $\beta$ transition for the DP3 peptide sequence. Here, $s$ is the integrated path length. The mechanism is described in terms of the path that contributes maximally to the global dynamics.

$\beta$-hairpin conformations for a designed peptide sequence. The coexistence of these folds results in a low-temperature peak in the heat capacity profile, which is reminiscent of solid-solid type transitions observed in atomic and molecular clusters,[91] as well as biomolecular switches.[97–99] Mutation of Tyrosine at position 12 to a Serine alters the landscape topography significantly, and lifts the degeneracy in agreement with the experiment of Araki and Tamura.[40]. As expected, the mutant sequence no longer exhibits any thermodynamic or kinetic signatures of competition between the $\alpha$-helix and $\beta$-hairpin conformations. The loss of degeneracy is traced to the overall destabilisation of the $\beta$-hairpin conformation, which results from the disruption of key hydrogen-bonding interactions upon mutation. Interestingly, mutation does not seem to alter the microscopic details of the transition mechanism between the $\alpha$-helix and the $\beta$-hairpin conformations substantially, suggesting that there could be generic features of this transformation that are conserved across different sequences.

## Acknowledgement

The authors thank Dr Sandeep Somani, Dr Chris Whittleston and Dr Joanne Carr for helpful discussions. We thank the ERC and the EPSRC for financial support. D.C. acknowledges the Cambridge Commonwealth, European, and International Trusts for providing a scholarship while he was a PhD student at the University of Cambridge, and Christ's College for a travel bursary. The authors declare no competing financial interests.

## References

1 C. B. Anfinsen, *Science*, 1973, **181**, 223–230.
2 X. I. Ambroggio and B. Kuhlman, *Curr. Opin. Struct. Biol.*, 2006, **16**, 525–530.
3 M. Gerstein and N. Echols, *Curr. Opin. Struct. Biol.*, 2004, **8**, 14–19.
4 J. H. Ha and S. N. Loh, *Chem. Eur. J.*, 2012, **18**, 7984–7999.
5 M. V. Golynskiy, M. S. Koay, J. L. Vinkenborg and M. Merkx, *ChemBioChem*, 2011, **12**, 353–361.
6 P. N. Bryan and J. Orban, *Curr. Opin. Struct. Biol.*, 2010, **20**, 482–488.
7 M. Menzei, *Protein Eng.*, 1998, **11**, 411–414.
8 L. Zhong and W. C. Johnson, *Proc. Natl. Acad. Sci. USA*, 1992, **89**, 4462–4465.
9 B. I. Cohen, S. R. Presnell and F. E. Cohen, *Prot. Sci.*, 1993, **2**, 2134–2145.
10 R. G. M. Mutter, U. Buttkus and K. H. Altmann, *Angew. Chem. Intl. Ed.*, 1991, **30**, 1514–1516.
11 W. Kabsch and C. Sander, *Proc. Natl. Acad. Sci. USA*, 1984, **81**, 1075–1078.
12 D. L. Minor and P. S. Kim, *Nature*, 1996, **380**, 730–734.
13 C. G. Roessler, B. M. Hall, W. J. Anderson, W. M. Ingram, S. A. Roberts, W. R. Montfort and M. H. J. Cordes, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 2343–2348.
14 G. D. Rose and T. P. Creamer, *Proteins*, 1994, **19**, 1–3.
15 S. Dalal, S. Balasubramanian and L. Regan, *Nat. Struct. Biol.*, 1997, **4**, 548–552.
16 S. Dalal, S. Balasubramanian and L. Regan, *Fold. Des.*, 1997, **2**, R71–R79.
17 P. A. Alexander, Y. He, Y. Chen, J. Orban and P. N. Bryan, *Proc. Natl. Acad. Sci. USA*, 2007, **104**, 11963–11968.
18 Y. He, Y. Chen, P. Alexander, P. N. Bryan and J. Orban, *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 14412–14417.
19 X. I. Ambroggio and B. Kuhlman, *J. Am. Chem. Soc.*, 2006, **128**, 1154–1161.
20 M. J. Pandya, E. Cerasoli, A. Joseph, R. G. Stoneman, E. Waite and D. N. Woolfson, *J. Am. Chem. Soc.*, 2004, **126**, 17016–17024.
21 B. D. Allen and S. L. Mayo, *J. Comput. Chem.*, 2010, **31**, 904–916.
22 C. Soto, *Nat. Rev. NeuroSci.*, 2003, **4**, 49–60.
23 M. Gross, *Curr. Protein Pept. Sci.*, 2000, **1**, 339–347.
24 S. B. Prusiner, *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 13363–13383.
25 K. M. Pan, M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. Huang, R. J. Fletterick and F. E. Cohen, *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 10962–10966.
26 J. Li, M. Shinjo, Y. Matsumara, M. Morita, D. Baker, M. Ikeguchi and H. Kihara, *Biochemistry*, 2007, **46**, 5072–5082.
27 D. Hamada, S.-I. Segawa and Y. Goto, *Nat. Struct. Biol.*, 1996, **3**, 868–873.
28 W. Yassine, N. Taib, S. Federman, A. Milochau, S. Castano, W. Sbi, C. Manigand, M. Laguerre, B. Desbat, R. Oda and J. Lang, *Biochim. Biophys. Acta - Biomembr.*, 2009, **1788**, 1722–1730.
29 Y. Takahashi, A. Ueno and H. Mihara, *Structure*, 2000, **8**, 915–925.
30 F. Simona, G. Tiana, R. A. Broglia and G. Colombo, *J. Mol. Graph. Model.*, 2004,

**23**, 263–273.
31 F. Ding, J. M. Borreguero, S. V. Buldyrey, H. E. Stanley and N. V. Dokholyan, *Proteins Struct. Funct. Genet.*, 2003, **53**, 220–228.
32 Z. Qin and M. J. Buehler, *Phys. Rev. Lett.*, 2010, **104**, 198304.
33 S. Singh, C. C. Chiu, A. S. Reddy and J. J. de Pablo, *J. Chem. Phys.*, 2013, **138**, 155101.
34 K. Ikeda and J. Higo, *Protein Sci.*, 2003, **12**, 2542–2548.
35 J. E. Straub, J. Guevara, S. Huo and J. P. Lee, *Acc. Chem. Res.*, 2002, **35**, 473–481.
36 V. Ovchinnikov and M. Karplus, *J. Chem. Phys.*, 2014, **140**, 175103.
37 N. A. Bernhardt, W. Xi, W. Wang and U. H. E. Hansmann, *J. Chem. Theory Comput.*, 2016, **12**, 5656–5666.
38 M. Andrec, A. K. Felts, E. Gallicchio and R. M. Levy, *Proc. Natl. Acad. Sci. USA*, 2004, **102**, 6801–6806.
39 H.-L. Chang, C.-J. Chen, H. Okumura and C.-K. Hu, *J. Comput. Chem.*, 2014, **35**, 1430–1437.
40 M. Araki and A. Tamura, *Proteins*, 2007, **66**, 860–868.
41 S. J. Demarest, R. Fairman and D. P. Rayleigh, *J. Mol. Biol.*, 1998, **283**, 279–281.
42 S. G. Itoh and Y. Okamoto, *Mol. Phys.*, 2007, **33**, 83–89.
43 S. G. Itoh, A. Tamura and Y. Okamoto, *J. Chem. Theory Comput.*, 2010, **6**, 979–983.
44 H. Okumura and S. G. Itoh, *Phys. Chem. Chem. Phys.*, 2013, **15**, 13852–13861.
45 C.-Y. Lin, N.-Y. Chen and C. Y. Mou, *Biophys J*, 2012, **103**, 99–108.
46 D. J. Wales, *Mol. Phys.*, 2002, **100**, 3285–3305.
47 D. J. Wales, *Mol. Phys.*, 2004, **102**, 891–908.
48 V. Ovchinnikov, K. Nam and M. Karplus, *J. Phys. Chem. B*, 2016, **120**, 8457–8472.
49 L. Xie and Z.-N. Chen, *Mol. Phys.*, 2016, **114**, 2424–2431.
50 J. A. Joseph, D. Chakraborty and D. J. Wales, *J. Chem. Theory Comput.*, 2018, **15**, 731–742.
51 Schrödinger, LLC, *The PyMOL Molecular Graphics System, Version 1.4.1*, 2010.
52 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins*, 2006, **65**, 712–725.
53 E. Malolepsza, B. Strodel, M. Khalili, S. Trygubenko, S. N. Fejer and D. J. Wales, *J. Comput. Chem.*, 2010, **31**, 1402–1409.
54 A. Onufriev, D. Bashford and D. A. Case, *Proteins*, 2004, **55**, 383–394.
55 A. Onufriev, D. Bashford and D. A. Case, *J. Phys. Chem. B.*, 2000, **104**, 3712–3720.
56 D. A. Case, T. A. Darden, T. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz and I. Kolossváry, *AMBER 12*, http://ambermd.org/, 2012.
57 D. J. Wales, *GMIN: A program for finding global minima and calculating thermodynamic properties from basin-sampling.*, http://www-wales.ch.cam.ac.uk/software.html.
58 J. M. Carr and D. J. Wales, *J. Phys. Chem. B*, 2008, **112**, 8760–8769.
59 D. E. Evans and D. J. Wales, *J. Chem. Phys.*, 2004, **121**, 1080–1090.
60 D. Chakraborty, R. Collepardo-Guevara and D. J. Wales, *J. Am. Chem. Soc.*, 2014, **136**, 18052–18061.
61 J. D. Farrell, C. Lines, J. J. Shepherd, D. Chakrabarti, M. A. Miller and D. J. Wales, *Soft Matter*, 2013, **9**, 5407–5416.
62 D. Chakraborty, N. Sengupta and D. J. Wales, *J. Phys. Chem. B*, 2016, **120**, 4331–4340.
63 J. Hernández-Rojas, D. Chakrabarti and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2016, **18**, 26579–26585.
64 J. N. Murrell and K. J. Laidler, *Trans. Faraday Soc.*, 1968, **64**, 371–377.
65 D. J. Wales, *Energy Landscapes*, Cambridge University Press, U.K., 2003.
66 S. A. Trygubenko and D. J. Wales, *J. Chem. Phys.*, 2004, **120**, 2082–2094.
67 G. Henkelman, B. P. Uberuaga and H. Jònsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.
68 G. Henkelman and H. Jònsson, *J. Chem. Phys.*, 1999, **111**, 7010–7022.
69 D. J. Wales and J. M. Carr, *J. Chem. Theory. Comput.*, 2012, **8**, 5020–5034.
70 L. J. Munro and D. J. Wales, *Phys. Rev. B.*, 1999, **59**, 3969–3980.
71 D. J. Wales, *OPTIM: A program for optimising geometries and calculating pathways*, http://www-wales.ch.cam.ac.uk/software.html.
72 D. A. Case, T. A. Darden, T. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Goetz and I. Kolossváry, *AMBER 9*, http://ambermd.org/, 2006.
73 D. Liu and J. Nocedal, *Math. Program.*, 1989, **45**, 503–528.
74 B. Strodel, C. W. Whittleston and D. J. Wales, *J. Am. Chem. Soc.*, 2007, **129**, 16005–16014.
75 J. M. Carr and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2009, **11**, 3341–3354.
76 E. W. Dijkstra, *Numer. Math.*, 1959, **1**, 269–271.
77 M. R. Hoare and J. J. McInnes, *Faraday Discuss. Chem. Soc.*, 1976, **61**, 12–24.
78 M. R. Hoare, in *Advances in Chemical Physics*, John Wiley and Sons, USA, 1979, vol. 40, pp. 49–129.
79 B. Strodel and D. J. Wales, *Chem. Phys. Lett.*, 2008, **466**, 105–115.
80 D. J. Wales, *J. Chem. Phys.*, 2009, **130**, 204111(1)–204111(7).
81 D. J. Wales and P. Salamon, *Proc. Natl. Acad. Sci. USA*, 2014, **111**, 617–622.
82 V. K. de Souza, J. D. Stevenson, S. P. Niblett, J. D. Farrell and D. J. Wales, *J. Chem. Phys.*, 2017, **146**, 124103.
83 O. M. Becker and M. Karplus, *J. Chem. Phys.*, 1997, **106**, 1495–1517.
84 S. V. Krivov and M. Karplus, *J. Chem. Phys.*, 2002, **117**, 10894–10903.
85 D. J. Wales, M. A. Miller and T. R. Walsh, *Nature*, 1998, **394**, 758–760.
86 D. A. Evans and D. J. Wales, *J. Chem. Phys.*, 2003, **118**, 3891–3897.
87 S. V. Krivov and M. Karplus, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14766–14770.
88 S. Somani and D. J. Wales, *J. Chem. Phys.*, 2013, **139**, 121909–1–121909–17.
89 D. J. Wales and T. Head-Gordon, *J. Phys. Chem. B.*, 2012, **116**, 8394–8411.
90 D. J. Wales, *J. Chem. Phys.*, 2015, **142**, 130901.
91 J. P. K. Doye, M. A. Miller and D. J. Wales, *J. Chem. Phys.*, 1999, **110**, 6896–6906.
92 Y. Levy, S. S. Cho, T. Shen, J. N. Onuchic and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 2373–2378.
93 K. Roder and D. J. Wales, *J. Phys. Chem B*, 2018, **122**, 10989–10995.
94 N. A. Bernhardt and U. H. E. Hansmann, *J. Phys. Chem. B*, 2018, **122**, 1600–1607.
95 S. Li, B. Xiong, Y. Xu, T. Lu, X. Luo, C. Luo, J. Shen, K. Chen, M. Zheng and H. Jiang, *J. Chem. Theory Comput.*, 2014, **10**, 2255–2264.
96 J. B. GC, Y. R. Bhandari, B. S. Gerstman and P. P. Chapagain, *J. Phys. Chem. B*, 2014, **118**, 5101–5108.
97 D. Chakraborty and D. J. Wales, *J. Chem. Phys.*, 2019, **150**, 125101.
98 K. Roder and D. J. Wales, *J. Chem Theory Comput.*, 2017, **13**, 1468–1477.
99 D. Chakraborty and D. J. Wales, *J. Phys. Chem. Lett*, 2018, **9**, 229–241.